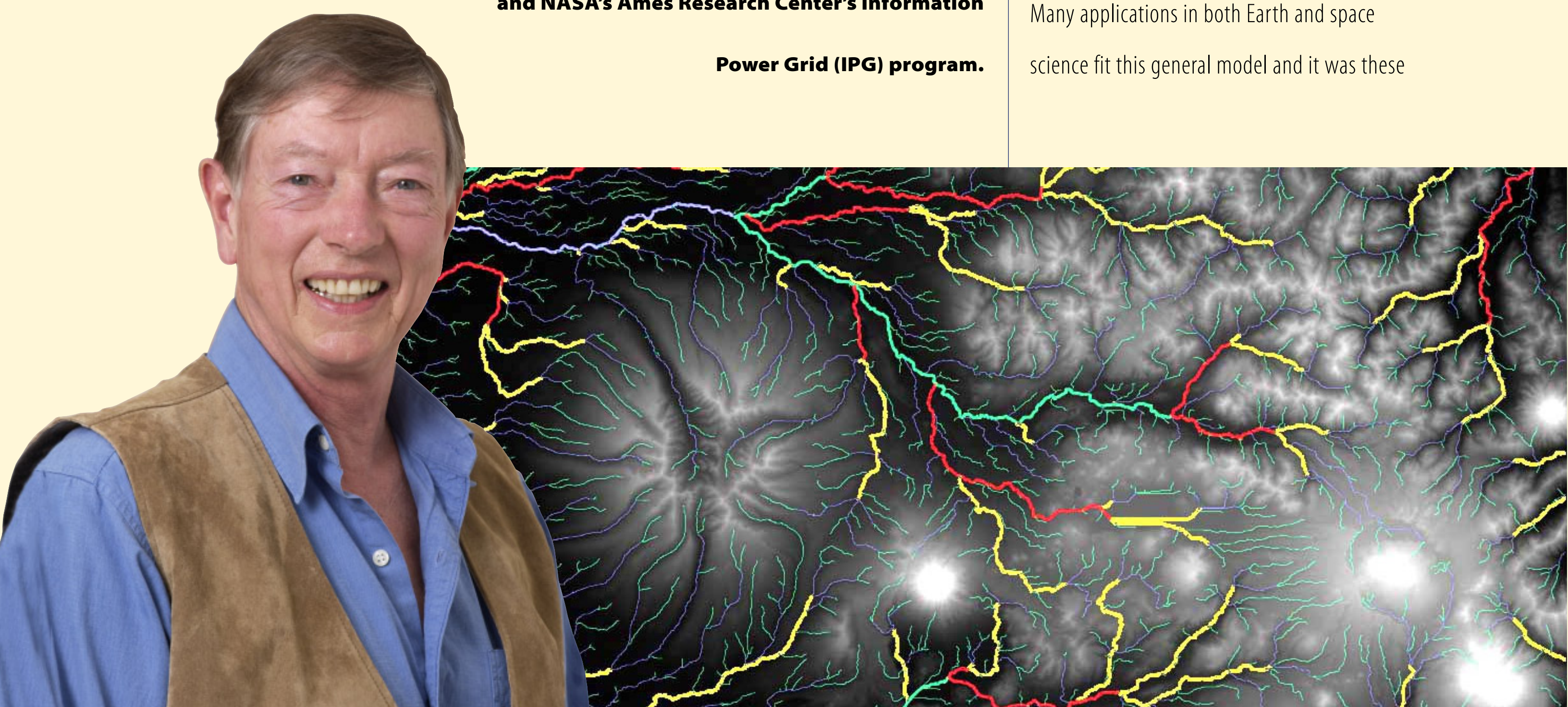
# Earth and Space Science Computing with the Information Power Grid

**"Grid computing" is simply the logical extension of "distributed computing," but for high-end machines and networks — distributed supercomputing, if you will. More formally, computational grids are persistent networked environments that integrate geographically distributed supercomputers, large databases, and high-end instruments. These resources are managed by diverse organizations in widespread locations, and are shared by researchers from many institutions. Recently the term "global grid" has become popular. Global grid is described as a collection of research and software infrastructure enabling grid computing. The community is self-organized and comes together under the title of Global Grid Forum. Leading practitioners have been the National Partnership for Advanced Computational Infrastructure (NPACI), the National Center for Supercomputing Applications (NCSA), and NASA's Ames Research Center's Information Power Grid (IPG) program.**

These grids have become sufficiently mature to be of interest for use in real science, perhaps in ways not possible without the one-sign-up, shared-facility approach that is at the heart of the grid access facilities. For example, we are implementing some aspects of the National Virtual Observatory (NVO) on top of the grid infrastructure. The promise is that by so doing, we can build simple interface mechanisms that allow even the small and thinly connected user to reach all of the nationally (and globally) distributed astronomical data archives, bring them to computational centers for large-scale computing, and return the results to the user. With this vision in place, not only do the data and computational cycles live on the grid, the software lives there as well. When this is reality, the individual researcher needs to have little else than a workstation and a research plan to do big-city science.

Many applications in both Earth and space science fit this general model and it was these



opportunities that led Ames Research Center to enlist JPL in putting together a set of high-impact science studies that could tap the grid's power. To pursue these ideas, Dave Curkendall is leading a new effort to bring three JPL applications to the grid — two in Earth science and one an extension of the NVO work. We are confident that this union of scientists and information technologists will prove a productive one. Indeed, it is opportunities like these that have led to the formation of the Earth System Information Technology Office. This office will seek to bring better focus and coordination to the Section's efforts to bring world-class information technology to world-class science.

*Left:* A view of the sky near the Galactic Center as seen in the infrared. From a mosaic of 2Mass images constructed by the NVO Montage Project. *Opposite page:* SRTM digital elevation map of Kamchatka in eastern Siberia with derived stream network overlay.

## Science Applications

### MULTISCALE OCEAN SIMULATION

This will be carried out in collaboration with Dr. Yi Chao, whose objectives are to do regional ocean modeling of the central California coast using a simulation cell size of 5 kilometers. Because the region to be modeled is sensibly influenced by the ocean's dynamics throughout the Pacific basin, a much larger simulation must be executed simultaneously. This can be of lower resolution — 15 kilometers for the whole West Coast area and then the entire Pacific domain at 50 kilometers. Since the IPG can support simultaneous scheduling and execution over multiple machines, we are devising a model where each simulation operates on a separate machine with the periodic exchange of boundary conditions over the networks.

### CONTINENTAL STREAM NETWORKS FROM SRTM TOPOGRAPHIC DATA

In collaboration with Dr. Eric Fielding and Dr. Ron Blom, we are seeking to exploit the emerging Shuttle Radar Topography Mission (SRTM) data to derive a detailed map of the stream networks, slopes, and drainage area on a continental basis. Stream and drainage networks are used in a variety of fields, but they are a primary data set for surface water hydrology. Hydrologists combine information about rainfall or river flow with the drainage network to determine where and how much water will arrive downstream. This is critical for evaluation of flood hazards and water resources. We seek to derive these networks globally and offer results over the Web as an overlay to the digital elevation model itself.

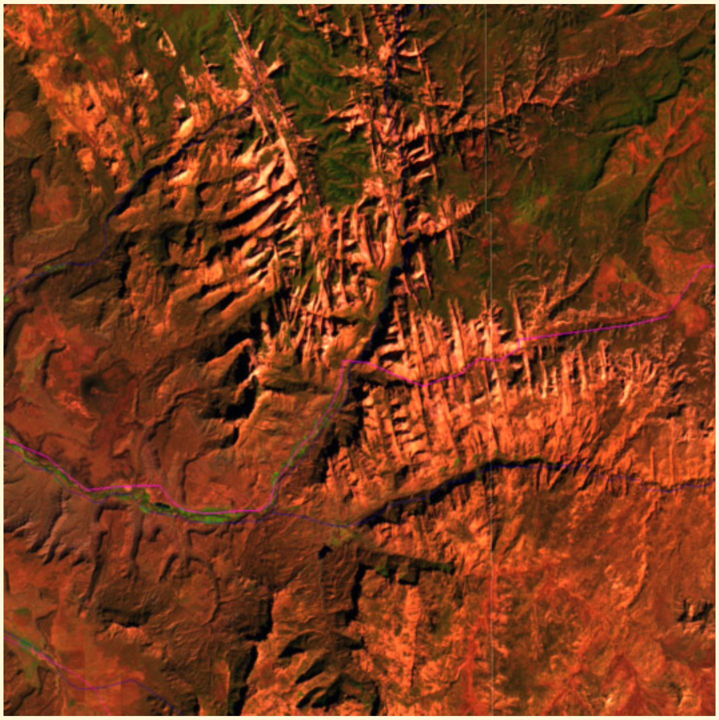
### A BASIC "ALL-SKY" MULTIBAND PLATE SET FOR THE NVO

This effort will be carried out in collaboration with the NASA Earth Science Technology Office Computational Technologies investigation, "High Performance Cornerstone Technologies for the NVO." One of the goals for the NVO is to develop a custom mosaic tool operating on the grid accessed through a Web portal (see *http://yoursky.jpl.nasa.gov* for a prototype of what will be called Montage). To provide a convenient browse tool, we are constructing a medium-resolution sky image (*http:/yoursky.jpl.nasa.gov/dsvo/*). Under the IPG program, we will extend this browse image to the full inherent resolution of the data — 1 arcsecond. The intent is to provide an all-sky set of reference plates of both optical and infrared data, processed with science-grade resampling and background removal algorithms. A Montage user will have a choice — to quickly extract a reference plate, or to tap the full Montage power with appeal to the original data and to the versatile custom mosaic tools.


## Contact

**Dr. Dave Curkendall**
**Jet Propulsion Laboratory/MS 126-103**
**4800 Oak Grove Drive**
**Pasadena, California 91109-8099**
*E-mail:* **dwc@jpl.nasa.gov**
*Tel:* **818-354-2406**

# A New Dimension in Earth Imaging: MapUS and Beyond



**NASA's Earth Science Enterprise seeks to study Earth as an interconnected system of atmosphere, oceans, continents, and life to "understand and protect our home planet." For this purpose, NASA has deployed a network of space and suborbital instrument platforms, which acquire a rapidly increasing volume of data. Supercomputing assets are needed to manage the incoming data stream and rapidly generate and distribute, via the Internet, custom data bundles to a multitude of users.**

*Above:* Spectacular geology of Zion National Park.

Much of the remote-imaging data is so rich in content that it defies any single purpose. It has enormous scientific and economic value; it can help generate unique insights in social and political applications; and it can greatly enhance science education. The challenge is to develop tools to help researchers, scientists, educators, and the interested public access and process these data via Internet links to remote supercomputers.

The Geographical Information Systems (GIS) industry, with NASA as an important contributor, has rallied behind the OpenGIS initiative to stimulate development of interoperable applications to mine the wealth of information from remote sensing. Sharing Earth image products has been one of the first objectives of OpenGIS. This objective was first realized with the introduction of the Web Mapping Server (WMS) specification.

MapUS *(http://mapus.jpl.nasa.gov)*, one of the early WMS servers built by Mr. Plesea in 1999, offers easy access to a 1-arcsecond (30-meter), seamless continental U.S. image mosaic. MapUS hosts an image mosaic built from 429 Landsat 5 images that date from 1992. It makes the mosaic available on the Internet both as a WMS service and as a stand-alone Web site. This mosaic preserves the six high-resolution spectral channels of Landsat 5. Through the Web interface, a user can combine any three channels to generate a color image. Other facilities provided by MapUS include progressive vector map overlay functions, realistic shadows derived from a U.S. elevation model, and orthogonal projection.

MapUS has generated more than a quarter million custom map images in the two years it has been available to the public, with peak access rates in excess of 20 maps per minute. The Web server was built as a prototype for delivering rapid Web access to custom mapping products through efficient use of supercomputing. As witnessed by frequent user comment, this goal has been achieved, with the map generation latency for even the most complex images being below three seconds. Plans are now in place to improve this even further.

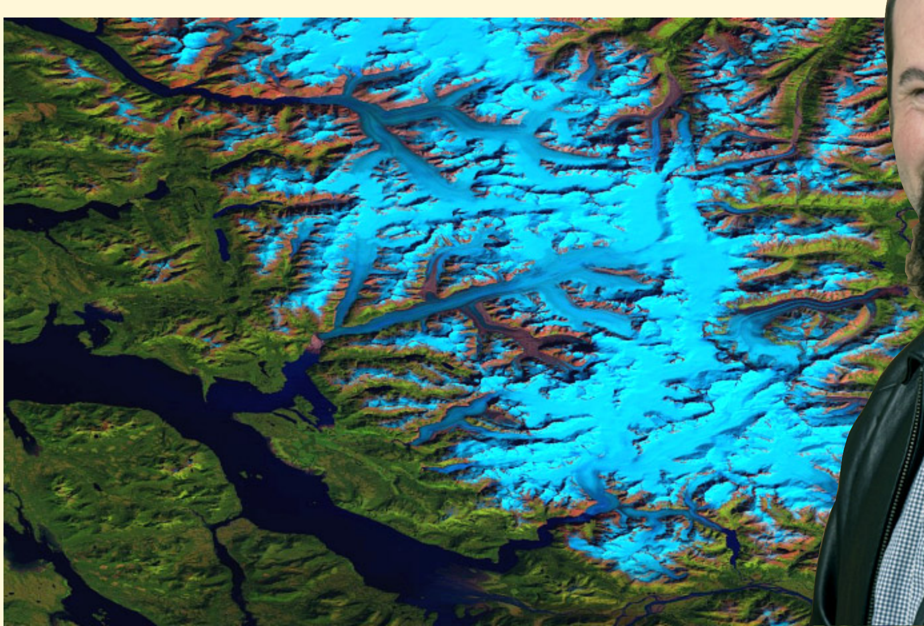
Encouraged by the success of MapUS, an ambitious new project has recently started, with the goal of assembling and providing Web access to a complete Earth landmass map image at 15-meter resolution. The complete Earth image will be nearly 100 times larger than the U.S. image, containing more than 1.5 trillion pixels. The images for this new mosaic, from Landsat 7, will represent a more current crop collected in 1999–2000.

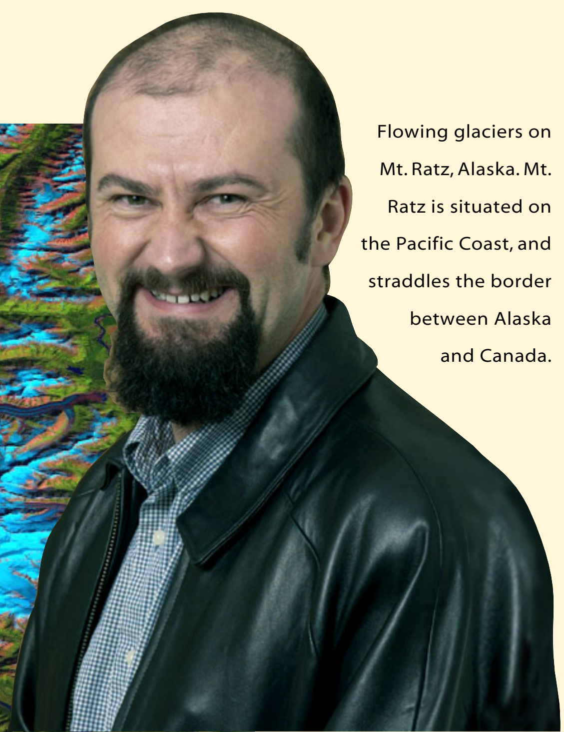
The vastness of this new dataset will stretch current technology limits. It will be possible to build and work with the mosaic only through proper application of supercomputers. In the area of disk storage alone, more than 10 terabytes will be required to hold the original data, intermediate processing steps, and the final mosaic. To cope with this at a reasonable cost we will use a pioneering approach to create a Linux disk farm from inexpensive consumer components. This storage cluster will closely resemble the Linux compute clusters known as Beowulf. The final cost for 10 terabytes of storage, based on 2003 prices, will be below $25,000. The required computation, estimated at 100,000 CPU-hours, will be provided by SGI Origin supercomputers at various locations within the NASA Information Power Grid.

Together with the global elevation model now being created by the joint NASA–National Imagery and Mapping Agency Shuttle Radar Topography Mission (SRTM) project, this mosaic will constitute essential data for the study of Earth systems, and will help to advance science analysis and modeling with real data sources at a resolution and scale not previously achieved.

### Contact

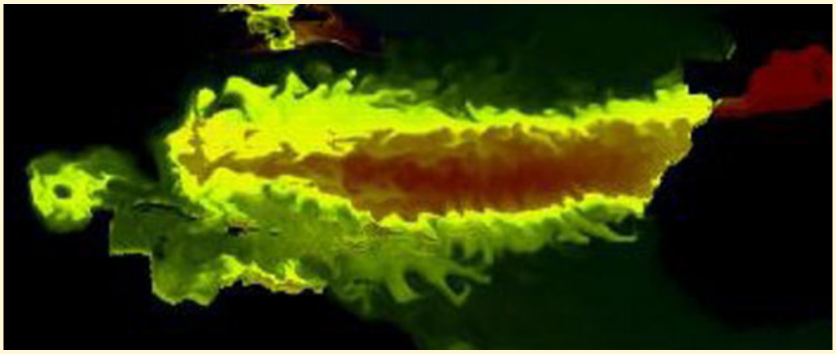
**Lucian Plesea**
**Jet Propulsion Laboratory/MS 126-104**
**4800 Oak Grove Drive**
**Pasadena, California 91109-8099**
*E-mail:* **Lucian.Plesea@jpl.nasa.gov**
*Tel:* **818-354-3928**





Flowing glaciers on Mt. Ratz, Alaska. Mt. Ratz is situated on the Pacific Coast, and straddles the border between Alaska and Canada.

# OurOcean: A Web Portal to Serve Near-Real-Time Coastal Ocean Data Products



3-D ocean salinity model data.

**The study of coastal ocean problems generally requires higher resolution and higher quality coastal data than the Earth Observing System (EOS) satellites can provide. Since the observations from satellites are either degraded or missing near the coastline, special algorithms are needed to refine the data in coastal regions. In addition, high-resolution, three-dimensional numerical models are crucial for estimating the state of the coastal ocean. For these reasons, a Web portal serving real-time, multidisciplinary coastal ocean data products will benefit many applications, such as weather forecasting, fishing, and long-term studies of the ocean/climate relationship.**

In collaboration with JPL oceanographer Dr. Yi Chao, Dr. Peggy Li and Dr. Joseph Jacob have developed a Web portal prototype called OurOcean *(http://ourocean.jpl.nasa.gov)*. OurOcean is an end-to-end system for data retrieval, data archiving, data processing, and data distribution, with a focus on the East Pacific Ocean. The purpose of this portal is to afford users easy access to ocean science data, run data assimilation models and visualize both data and models. Through OurOcean, users with minimal resources can access large datasets and interact with sophisticated ocean models.

Currently, OurOcean provides both real-time and retrospective analysis of remote sensing data and ocean model simulations in the Pacific Ocean, with a focus on the coastal ocean from Baja California to the Gulf of Alaska.

OurOcean serves the following datasets:

- Daily ocean surface wind datasets collected by the QuikScat scatterometer and stored at JPL's Physical Oceanographic DAAC Center.
- Twice-daily ocean surface wind from the Coupled Ocean Atmosphere Mesoscale Predication System (COAMPS) model stored at the U.S. Navy's Fleet Numerical Meteorology and Oceanography Center (FNMAC).
- A blended wind product produced at JPL.

- A three-dimensional ocean temperature, salinity, and current velocity dataset generated by the Regional Ocean Model System (ROMS).

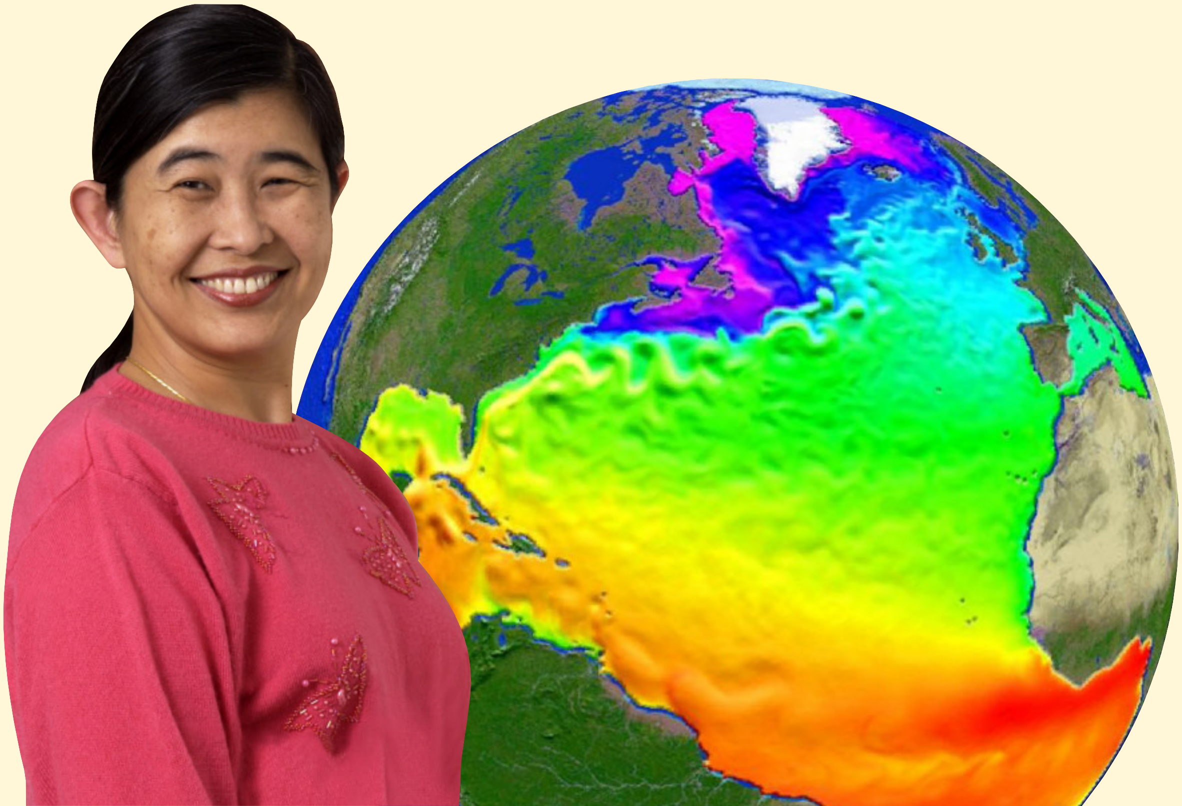
The Live Access Server (LAS) is used to serve data over the Web, and the Ferret Data Visualization and Analysis System is used to render visual products requested by the user. OurOcean users work with an interactive Web interface that allows them to select a subset of a larger dataset by clicking and dragging on a zoomable map. They can choose from a variety of outputs, such as a GIF or PostScript image, a raw data file, or a NetCDF file. A user can also compare the difference between any two variables.

The OurOcean server downloads the QuikScat and COAMP datasets daily from their distributed archives, re-grids and re-formats them, runs a blending program to generate the blended wind dataset and puts all three datasets on the Web. The blended wind data feed into a three-dimensional model running on a remote supercomputer that generates real-time, three-dimensional ocean datasets, available from OurOcean. OurOcean provides a unified interface to diverse users for accessing multidisciplinary ocean datasets focusing on a specific region of interest.

In the future, OurOcean may be expanded to serve on-demand modeling datasets. A user will be able to select a region of interest, data resolution, simulation duration, and boundary conditions of the model via the Web interface and submit a request to the Web server. The server will then configure the model, schedule remote supercomputing resources to run the model, manage the data assimilation and data transfer, monitor the model execution, and deliver the output to the OurOcean Web server for the user to retrieve. Eventually, OurOcean will serve as a "one-stop shop" for very large distributed datasets and easy interaction with complex three-dimensional models.

## Contact

**Dr. Peggy Li**
**Jet Propulsion Laboratory/MS 126-104**
**4800 Oak Grove Drive**
**Pasadena, California 91109-8099**
*E-mail:* P.P.Li@jpl.nasa.gov
*Tel:* 818-354-1341



Ocean surface temperature model data.